

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 August 2002 (29.08.2002)

PCT

(10) International Publication Number  
**WO 02/066624 A2**

(51) International Patent Classification<sup>7</sup>: C12N 9/00

(21) International Application Number: PCT/US02/04863

(22) International Filing Date: 19 February 2002 (19.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/270,320 20 February 2001 (20.02.2001) US

(71) Applicant: **LEXICON GENETICS INCORPORATED**  
[US/US]; 4000 Research Forest Drive, The Woodlands, TX  
77381 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors: **HU, Yi**; 333 Holly Creek Ct. #203, The Woodlands, TX 77381 (US). **NEPOMNICHY, Boris**; 905 Cypress Station Dr. #1515, Houston, TX 77909 (US).

**Published:**

— without international search report and to be republished upon receipt of that report

(74) Agents: **ISHIMOTO, Lance, K.** et al.; LEXICON GENETICS INCORPORATED, 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN PROTEASE AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

WO 02/066624 A2

NOVEL HUMAN PROTEASE AND  
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.  
5 Provisional Application Number 60/270,320, which was filed on  
February 20, 2001, and is herein incorporated by reference in  
its entirety.

1. INTRODUCTION

10 The present invention relates to the discovery,  
identification, and characterization of novel human  
polynucleotides encoding a protein sharing sequence similarity  
with mammalian proteases. The invention encompasses the  
described polynucleotides, host cell expression systems, the  
15 encoded protein, fusion proteins, polypeptides and peptides,  
antibodies to the encoded proteins and peptides, and  
genetically engineered animals that either lack or overexpress  
the disclosed polynucleotides, antagonists and agonists of the  
proteins, and other compounds that modulate the expression or  
20 activity of the proteins encoded by the disclosed  
polynucleotides, which can be used for diagnosis, drug  
screening, clinical trial monitoring, the treatment of  
diseases and disorders, and cosmetic or nutraceutical  
applications.

25

2. BACKGROUND OF THE INVENTION

Proteases cleave protein substrates as part of  
degradation, maturation, and secretory pathways within the  
body. Proteases have been associated with, *inter alia*,  
30 regulating development, modulating cellular processes,  
fertility, and infectious disease. Proteases are therefore  
attractive drug targets.

## 3. SUMMARY OF THE INVENTION

The present invention relates to the discovery, identification, and characterization of nucleotides that encode a novel human protein, and the corresponding amino acid sequence of this protein. The novel human protein (NHP) described for the first time herein shares structural similarity with animal proteases, and particularly matrix metalloproteases, zinc dependent metalloproteases, and bone morphogenetic protein.

The novel human nucleic acid (cDNA) sequences described herein encode a protein/open reading frame (ORF) of 436 amino acids in length (SEQ ID NO:2).

The invention also encompasses agonists and antagonists of the described NHP, including small molecules, large molecules, mutant NHPs, or portions thereof, that compete with native NHP, peptides, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHP (e.g., antisense and ribozyme molecules, and open reading frame or regulatory sequence replacement constructs) or to enhance the expression of the described NHP (e.g., expression constructs that place the described polynucleotide under the control of a strong promoter system), and transgenic animals that express a NHP sequence, or "knock-outs" (which can be conditional) that do not express a functional NHP. Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cell ("ES cell") lines that contain gene trap mutations in a murine homolog of the described NHP. When the unique NHP sequences described in SEQ ID NOS:1-3 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene, as well as a method of assigning function to previously unknown genes. In addition, animals in which the

unique NHP sequences described in SEQ ID NOS:1-3 are "knocked-out" provide a unique source in which to elicit antibodies to homologous and orthologous proteins, which would have been previously viewed by the immune system as "self" and therefore would have failed to elicit significant antibody responses.

Additionally, the unique NHP sequences described in SEQ ID NOS:1-3 are useful for the identification of protein coding sequences, and mapping a unique gene to a particular chromosome. These sequences identify biologically verified exon splice junctions, as opposed to splice junctions that may have been bioinformatically predicted from genomic sequence alone. The sequences of the present invention are also useful as additional DNA markers for restriction fragment length polymorphism (RFLP) analysis, and in forensic biology.

Further, the present invention also relates to processes for identifying compounds that modulate, *i.e.*, act as agonists or antagonists of, NHP expression and/or NHP activity that utilize purified preparations of the described NHP and/or NHP products, or cells expressing the same. Such compounds can be used as therapeutic agents for the treatment of any of a wide variety of symptoms associated with biological disorders or imbalances.

#### 4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing provides sequences encoding the described NHP amino acid sequence. SEQ ID NO:3 describes a NHP ORF and flanking regions.

#### 5. DETAILED DESCRIPTION OF THE INVENTION

The NHP described for the first time herein is a novel protein that can be expressed in, *inter alia*, human lymph node, mammary gland, and fetal kidney cells.

The described sequences were compiled from cDNAs prepared and isolated from human lymph node, mammary gland, and brain mRNAs (Edge Biosystems, Gaithersburg, MD, Clontech, Palo Alto, CA).

5       The present invention encompasses the nucleotides presented in the Sequence Listing, host cells expressing such nucleotides, the expression products of such nucleotides, and: (a) nucleotides that encode mammalian homologs of the described polynucleotides, including the specifically  
10 described NHP, and the NHP products; (b) nucleotides that encode one or more portions of the NHP that correspond to functional domains, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, the novel regions of any active domain(s); (c) isolated  
15 nucleotides that encode mutant versions, engineered or naturally occurring, of the described NHP in which all or a part of at least one domain is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, soluble proteins and peptides  
20 in which all or a portion of the signal sequence is deleted; (d) nucleotides that encode chimeric fusion proteins containing all or a portion of a coding region of the NHP, or one of its domains (e.g., a receptor or ligand binding domain, accessory protein/self-association domain, etc.) fused to  
25 another peptide or polypeptide; or (e) therapeutic or diagnostic derivatives of the described polynucleotides, such as oligonucleotides, antisense polynucleotides, ribozymes, dsRNA, or gene therapy constructs comprising a sequence first disclosed in the Sequence Listing.

30       As discussed above, the present invention includes the human DNA sequences presented in the Sequence Listing (and vectors comprising the same), and additionally contemplates

any nucleotide sequence encoding a contiguous NHP open reading frame (ORF) that hybridizes to a complement of a DNA sequence presented in the Sequence Listing under highly stringent conditions, e.g., hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1x SSC/0.1% SDS at 68°C (Ausubel et al., eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., N.Y., at p. 2.10.3) and encodes a functionally equivalent expression product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of a DNA sequence that encodes and expresses an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, e.g., washing in 0.2x SSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet still encode a functionally equivalent NHP product. Functional equivalents of a NHP include naturally occurring NHPs present in other species, and mutant NHPs, whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent No. 5,837,458). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides encoding a NHP ORF, or its functional equivalent, encoded by a polynucleotide sequence that is about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the nucleotide sequences of the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package, as described herein, using standard default settings).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore

the complements of, the described NHP nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described herein. In instances where the nucleic acid molecules are deoxyoligonucleotides ("DNA

5 oligos"), such molecules are generally about 16 to about 100 bases long, or about 20 to about 80 bases long, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence Listing. Such  
10 oligonucleotides can be used in conjunction with the polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc.

Alternatively, such NHP oligonucleotides can be used as hybridization probes for screening libraries, and assessing  
15 gene expression patterns (particularly using a microarray or high-throughput "chip" format). Additionally, a series of the described NHP oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NHP sequences. An oligonucleotide or polynucleotide  
20 sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS:1-3 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or  
25 polycrystalline substrates, etc.). Of particular note are spatially addressable arrays (i.e., gene chips, microtiter plates, etc.) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers present on the spatially addressable  
30 array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS:1-3, or an amino acid sequence encoded thereby. Methods

for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon, are disclosed in, *inter alia*, U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405, the disclosures of which are herein incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-3 can be used to identify and characterize the temporal and tissue specific expression of a gene. These addressable arrays incorporate oligonucleotide sequences of sufficient length to confer the required specificity, yet be within the limitations of the production technology. The length of these probes is usually within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60 nucleotides, and more preferably 25 nucleotides, from the sequences first disclosed in SEQ ID NOS:1-3.

For example, a series of the described oligonucleotide sequences, or the complements thereof, can be used in chip format to represent all or a portion of the described sequences. The oligonucleotides, typically between about 16 to about 40 (or any whole number within the stated range) nucleotides in length, can partially overlap each other, and/or the sequence may be represented using oligonucleotides that do not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 8 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences can begin at any nucleotide present within a sequence in the Sequence Listing, and proceed in



either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of broad patterns of genetic activity, providing new understanding of  
5 gene functions, and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-3 provides detailed information about transcriptional changes involved in a specific pathway,  
10 potentially leading to the identification of novel components, or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID NOS:1-3 can also be used in the identification, selection, and  
15 validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets, and recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the intended target of the drug. These unique sequences  
20 therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-3 can be utilized in microarrays, or other assay formats, to screen collections of genetic material from  
25 patients who have a particular medical condition. These investigations can also be carried out using the sequences first disclosed in SEQ ID NOS:1-3 *in silico*, and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

30 Thus the sequences first disclosed in SEQ ID NOS:1-3 can be used to identify mutations associated with a particular disease, and also in diagnostic and/or prognostic assays.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within a given region of the sequence, in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in SEQ ID NOS:1-3. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other specific oligonucleotide sequences, can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (e.g., the University of Wisconsin GCG sequence analysis package, SEQUENCHER 3.0, Gene Codes Corp., Ann Arbor, MI, etc.), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, e.g., to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as NHP antisense molecules, useful, for example, in NHP gene regulation and/or as antisense primers in amplification reactions of NHP nucleic acid sequences. With respect to NHP gene regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as

part of ribozyme and/or triple helix sequences that are also useful for NHP gene regulation.

Inhibitory antisense or double stranded oligonucleotides can additionally comprise at least one modified base moiety, which is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

The antisense oligonucleotide can also comprise at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide will comprise at least one modified phosphate backbone selected from the group including, but not limited to, a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate,

a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric  
5 oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res.  
10 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330). Alternatively, double stranded RNA can be used to disrupt the expression and function of a targeted NHP.

Oligonucleotides of the invention can be synthesized by  
15 standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides can be synthesized (Stein et al., 1988, Nucl. Acids Res. 16:3209), and methylphosphonate  
20 oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. USA 85:7448-7451), etc.

Low stringency conditions are well-known to those of skill in the art, and will vary predictably depending on the  
25 specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (and periodic updates thereof); and Ausubel et  
30 al., 1989, *supra*.

Alternatively, suitably labeled NHP nucleotide probes can be used to screen a human genomic library using appropriately

stringent conditions or by PCR. The identification and characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide polymorphisms, or coding single nucleotide polymorphisms), determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (e.g., splice acceptor and/or donor sites), etc., that can be used in diagnostics and pharmacogenomics.

For example, the present sequences can be used in restriction fragment length polymorphism (RFLP) analysis to identify specific individuals. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification (as generally described in U.S. Patent No. 5,272,057, incorporated herein by reference). In addition, the sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). Actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments.

Further, a NHP homolog can be isolated from nucleic acid from an organism of interest by performing PCR using two degenerate or "wobble" oligonucleotide primer pools designed

on the basis of amino acid sequences within the NHP products disclosed herein. The template for the reaction may be total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue  
5 known to express, or suspected of expressing, an allele of a NHP gene. The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For  
10 example, the amplified fragment can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate genomic clones via the screening of a genomic library.

PCR technology can also be used to isolate full length  
15 cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, one known to express, or suspected of expressing, a NHP gene, such as, for example, lymph node tissue). A reverse transcription (RT) reaction can be  
20 performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase reaction, the hybrid may be digested with RNase H, and second  
25 strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified fragment can be isolated. For a review of cloning strategies that can be used, see, *e.g.*, Sambrook et al., 1989, *supra*.

A cDNA encoding a mutant NHP sequence can be isolated,  
30 for example, by using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known to express,

or suspected of expressing, a NHP, in an individual putatively carrying a mutant NHP allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes  
5 specifically to the 5' end of the normal sequence. Using these two primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well-known to those of skill in the art. By comparing the DNA sequence of the mutant NHP  
10 allele to that of a corresponding normal NHP allele, the mutation(s) responsible for the loss or alteration of function of the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of carrying, or  
15 known to carry, a mutant NHP allele (e.g., a person manifesting a NHP-associated phenotype such as, for example, obesity, high blood pressure, connective tissue disorders, infertility, etc.), or a cDNA library can be constructed using RNA from a tissue known to express, or suspected of  
20 expressing, a mutant NHP allele. A normal NHP gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NHP allele in such libraries. Clones containing mutant NHP sequences can then be purified and subjected to sequence analysis according to  
25 methods well-known to those skilled in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known to express, or suspected of expressing, a mutant NHP allele in an individual suspected of carrying, or  
30 known to carry, such a mutant allele. In this manner, gene products made by the putatively mutant tissue can be expressed and screened using standard antibody screening techniques in

conjunction with antibodies raised against a normal NHP product, as described below (for screening techniques, see, for example, Harlow and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Additionally, screening can be accomplished by screening with labeled NHP fusion proteins, such as, for example, alkaline phosphatase-NHP or NHP-alkaline phosphatase fusion proteins. In cases where a NHP mutation results in an expression product with altered function (e.g., as a result of a missense or a frameshift mutation), polyclonal antibodies to the NHP are likely to cross-react with a corresponding mutant NHP expression product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well-known in the art.

The invention also encompasses: (a) DNA vectors that contain any of the foregoing NHP coding sequences and/or their complements (*i.e.*, antisense); (b) DNA expression vectors that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences (for example, baculovirus as described in U.S. Patent No. 5,869,336 herein incorporated by reference); (c) genetically engineered host cells that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host cells that express an endogenous NHP sequence under the control of an exogenously introduced regulatory element (*i.e.*, gene activation). As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators, and other



elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include, but are not limited to, the cytomegalovirus (hCMV) immediate early gene, regulatable, viral elements (particularly retroviral LTR promoters), the early or late promoters of SV40 or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast  $\alpha$ -mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of the NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP sequence (transcription factor inhibitors, antisense and ribozyme molecules, or open reading frame sequence or regulatory sequence replacement constructs), or promote the expression of a NHP (e.g., expression constructs in which NHP coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, etc.).

The NHP or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs, or inappropriately expressed NHPs, for the diagnosis of disease. The NHP protein or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of the NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such

systems allow not only for the identification of compounds that bind to the endogenous receptor for the NHP, but can also identify compounds that trigger NHP-mediated activities or pathways.

5        Finally, the NHP products can be used as therapeutics. For example, soluble derivatives such as NHP peptides/domains corresponding to the NHP, NHP fusion protein products (especially NHP-Ig fusion proteins, *i.e.*, fusions of the NHP, or a domain of the NHP, to an IgFc), NHP antibodies and anti-  
10    idiotypic antibodies (including Fab fragments), antagonists or agonists (including compounds that modulate or act on downstream targets in a NHP-mediated pathway) can be used to directly treat diseases or disorders. For instance, the administration of an effective amount of a soluble NHP, a NHP-  
15    IgFc fusion protein, or an anti-idiotypic antibody (or its Fab) that mimics the NHP could activate or effectively antagonize the endogenous NHP receptor. Nucleotide constructs encoding such NHP products can be used to genetically engineer host cells to express such products *in vivo*; these genetically  
20    engineered cells function as "bioreactors" in the body delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. Nucleotide constructs encoding a functional NHP, mutant NHPs, as well as antisense and ribozyme molecules can also be used in "gene therapy"  
25    approaches for the modulation of NHP expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

Various aspects of the invention are described in greater detail in the subsections below.

30

### 5.1 THE NHP SEQUENCES

The cDNA sequence (SEQ ID NO:1) and the corresponding deduced amino acid sequence (SEQ ID NO:2) of the described NHP are presented in the Sequence Listing. The gene encoding the described NHP is apparently present on human chromosome 2 (see, for example, GENBANK Accession No. AC012307). Accordingly, the described sequences are additionally useful for mapping the coding regions of the human chromosome.

An additional application of the described novel human polynucleotide sequences is their use in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences using, for example, polynucleotide shuffling or related methodologies. Such approaches are described in U.S. Patent Nos. 5,830,721 and 5,837,458, which are herein incorporated by reference in their entirety.

NHP gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, birds, goats, and non-human primates, e.g., baboons, monkeys, and chimpanzees, may be used to generate NHP transgenic animals.

Any technique known in the art may be used to introduce a NHP transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Patent No. 4,873,191); retrovirus-mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci. USA 82:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell 56:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*,

1989, Cell 57:717-723); etc. For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. 115:171-229, which is incorporated by reference herein in its entirety.

5       The present invention provides for transgenic animals that carry a NHP transgene in all their cells, as well as animals that carry a transgene in some, but not all their cells, *i.e.*, mosaic animals or somatic cell transgenic animals. A transgene may be integrated as a single transgene,  
10 or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. A transgene may also be selectively introduced into and activated in a particular cell-type by following, for example, the teaching of Lasko *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89:6232-6236. The regulatory sequences required for  
15 such a cell-type specific activation will depend upon the particular cell-type of interest, and will be apparent to those of skill in the art.

When it is desired that a NHP transgene be integrated into the chromosomal site of the endogenous NHP gene, gene  
20 targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous NHP gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the  
25 nucleotide sequence of the endogenous NHP gene (*i.e.*, "knockout" animals). The present invention also provides for "knockin" animals. Knockin animals are those having a gene that the animal does not naturally have inserted in its genome. For example, when a human gene is used to replace its  
30 murine ortholog in the mouse. Such knockin animals are useful for the *in vivo* study, testing and validation of, *intra alia*,

human biotherapeutics, drug targets, as well as for compounds that are directed at the same.

The transgene can also be selectively introduced into a particular cell-type, thus inactivating the endogenous NHP gene in only that cell-type, by following, for example, the teaching of Gu et al., 1994, Science 265:103-106. The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell-type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant NHP gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of NHP gene-expressing tissue may also be evaluated immunocytochemically using antibodies specific for the NHP transgene product.

## 5.2 NHP AND NHP POLYPEPTIDES

NHP, NHP polypeptides, NHP peptide fragments, mutated, truncated, or deleted forms of the NHP, and/or NHP fusion proteins can be prepared for a variety of uses. These uses include, but are not limited to, the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products related to the NHP, and as reagents in assays for screening for compounds that can be used as pharmaceutical reagents useful in the therapeutic

treatment of mental, biological, or medical disorders and diseases. Given the similarity information and expression data, the described NHP can be targeted (by drugs, oligos, antibodies, etc.) in order to treat disease, or to  
5 therapeutically supplant or augment the efficacy of, for example, chemotherapeutic agents used in the treatment of cancer, or agents used to treat inflammatory disorders, arthritis, or infectious diseases. Because of their medical importance, metalloproteases similar to the described NHP have  
10 been studied by others, as exemplified in U.S. Patent No. 5,922,546, herein incorporated by reference, which further describes a variety of uses that are also applicable to the described NHP.

The Sequence Listing discloses the amino acid sequence  
15 encoded by the described NHP polynucleotides. The ORF encoding the NHP displays an initiator methionine in a DNA sequence context consistent with a translation initiation site, and a signal sequence, which can indicate that the described NHP can be secreted or membrane-associated.

20 The NHP amino acid sequences of the invention includes the amino acid sequence presented in the Sequence Listing, as well as analogues and derivatives thereof. Further, corresponding NHP homologues from other species are encompassed by the invention. In fact, any NHP encoded by the  
25 NHP nucleotide sequences described herein are within the scope of the invention, as are any novel polynucleotide sequences encoding all or any novel portion of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well-known, and, accordingly, each amino  
30 acid presented in the Sequence Listing is generically representative of the well-known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As

such, as contemplated herein, the amino acid sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell et al., eds., Scientific American Books, New York, N.Y., herein incorporated by reference), are generically representative of all the various permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NHP encoded by the presently described nucleotide sequences, as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of the NHP, the ability to effect an identical or complementary downstream pathway, or a change in cellular metabolism (e.g., proteolytic activity, ion flux, tyrosine phosphorylation, etc.). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences described herein, but that result in a silent change, thus producing a functionally equivalent expression product. Amino acid substitutions can be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

A variety of host-expression vector systems can be used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP peptide or polypeptide is thought to be a soluble or secreted molecule, the peptide or polypeptide can be recovered from the culture media. Such expression systems also encompass engineered host cells that express a NHP, or functional equivalent, *in situ*. Purification or enrichment of a NHP from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well-known to those skilled in the art. However, such engineered host cells themselves may be used in situations where it is important not only to retain the structural and functional characteristics of a NHP, but to assess biological activity, e.g., in certain drug screening assays.

The expression systems that may be used for purposes of the invention include, but are not limited to, microorganisms such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NHP nucleotide sequences; yeast (e.g., *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NHP nucleotide sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing NHP nucleotide sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing NHP nucleotide sequences; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing NHP nucleotide sequences and promoters derived from the genome of mammalian cells (e.g.,



metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NHP product being expressed. For example, when a large quantity of such a protein is to be produced for the generation of pharmaceutical compositions of or containing a NHP, or for raising antibodies to a NHP, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which a NHP coding sequence may be ligated individually into the vector in-frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke and Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Pharmacia or American Type Culture Collection) can also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target expression product can be released from the GST moiety.

In an exemplary insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign polynucleotide sequences. The virus grows in *Spodoptera frugiperda* cells. A NHP coding sequence can be cloned individually into a non-essential region (for example the polyhedrin gene) of the virus and placed under control of

an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted sequence is expressed (*e.g.*, see Smith *et al.*, 1983, *J. Virol.* 46:584; Smith, U.S. Patent No. 4,215,051).

10 In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the NHP nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric sequence may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NHP product in infected hosts (*e.g.*, see Logan and Shenk, 1984, *Proc. Natl. Acad. Sci. USA* 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NHP nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NHP gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a NHP coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, may be provided. Furthermore, the initiation codon should be in phase with the

reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter et al., 1987, Methods in Enzymol. 153:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the expression product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and expression products. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for the desired processing of the primary transcript, glycosylation, and phosphorylation of the expression product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, human cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the NHP sequences described herein can be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences,

transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then switched to a selective media.

5 The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci, which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer  
10 cell lines that express the NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the NHP product.

A number of selection systems may be used, including, but  
15 not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977, *Cell* 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, 1962, *Proc. Natl. Acad. Sci. USA* 48:2026), and adenine phosphoribosyltransferase (Lowy *et al.*, 1980, *Cell* 22:817)  
20 genes, which can be employed in *tk*<sup>-</sup>, *hgp*<sup>-</sup> or *aprt*<sup>-</sup> cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: *dhfr*, which confers resistance to methotrexate (Wigler *et al.*, 1980, *Proc. Natl. Acad. Sci. USA* 77:3567; O'Hare *et al.*, 1981, *Proc. Natl. Acad. Sci. USA* 78:1527); *gpt*, which confers resistance to  
25 mycophenolic acid (Mulligan and Berg, 1981, *Proc. Natl. Acad. Sci. USA* 78:2072); *neo*, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981, *J. Mol. Biol.* 150:1); and *hygro*, which confers resistance to  
30 hygromycin (Santerre *et al.*, 1984, *Gene* 30:147).

Alternatively, any fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being

expressed. An exemplary system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht et al., 1991, Proc. Natl. Acad. Sci. USA 88:8972-8976). In this system, the sequence of  
5 interest is subcloned into a vaccinia recombination plasmid such that the sequence's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto  $\text{Ni}^{2+}$ -nitriloacetic acid-agarose  
10 columns, and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct a NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation  
15 of a NHP to an antibody molecule or its Fab fragment could be used to target cells bearing a particular epitope. Attaching an appropriate signal sequence to a NHP would also transport a NHP to a desired location within the cell. Alternatively targeting of a NHP or its nucleic acid sequence might be  
20 achieved using liposome or lipid complex based delivery systems. Such technologies are described in "Liposomes: A Practical Approach", New, R.R.C., ed., Oxford University Press, N.Y., and in U.S. Patent Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures,  
25 which are herein incorporated by reference in their entirety. Additionally embodied are novel protein constructs engineered in such a way that they facilitate transport of a NHP to a target site or desired organ, where it crosses the cell membrane and/or the nucleus where the NHP can exert its  
30 functional activity. This goal may be achieved by coupling of a NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see

generally U.S. Provisional Patent Application Ser. Nos. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples of such transducing sequences), to facilitate passage across cellular membranes, and can optionally be engineered to include nuclear localization signals.

### 5.3 ANTIBODIES TO NHP PRODUCTS

Antibodies that specifically recognize one or more epitopes of a NHP, epitopes of conserved variants of a NHP, or peptide fragments of a NHP are also encompassed by the invention. Such antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')<sub>2</sub> fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

The antibodies of the invention may be used, for example, in the detection of a NHP in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of a NHP. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression and/or activity of a NHP expression product. Additionally, such antibodies can be used in conjunction with gene therapy to, for example, evaluate normal and/or engineered NHP-expressing cells prior to their introduction into a patient. Such antibodies may additionally be used in methods for the inhibition of abnormal NHP activity. Thus, such antibodies may be utilized as a part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with the NHP, a NHP peptide (e.g., one corresponding to a functional domain of a NHP), truncated NHP polypeptides (a NHP in which one or more domains have been deleted), functional equivalents of the NHP, or mutated variants of the NHP. Such host animals may include, but are not limited to, pigs, rabbits, mice, goats, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including, but not limited to, Freund's adjuvant (complete and incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, chitosan, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Alternatively, the immune response could be enhanced by combination and/or coupling with molecules such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin, or fragments thereof. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495-497; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, *Immunology Today* 4:72; Cole et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Such antibodies may

be of any immunoglobulin class, including IgG, IgM, IgE, IgA, and IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this  
5 the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, 1984, Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger *et al.*, 1984, Nature, 312:604-608; Takeda *et al.*, 1985, Nature, 314:452-454), by  
10 splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity, can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species,  
15 such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Such technologies are described in U.S. Patent Nos. 6,114,598, 6,075,181 and 5,877,397 and their respective disclosures, which are herein incorporated by reference in their entirety.  
20 Also encompassed by the present invention is the use of fully humanized monoclonal antibodies, as described in U.S. Patent No. 6,150,584 and respective disclosures, which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of  
25 single chain antibodies (U.S. Patent No. 4,946,778; Bird, 1988, Science 242:423-426; Huston *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward *et al.*, 1989, Nature 341:544-546) can be adapted to produce single chain antibodies against NHP expression products. Single chain antibodies are  
30 formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.



Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to:  $F(ab')_2$  fragments, which can be produced by pepsin digestion of an antibody molecule; and  
5 Fab fragments, which can be generated by reducing the disulfide bridges of  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, *Science*, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

10 Antibodies to a NHP can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" a given NHP, using techniques well-known to those skilled in the art (see, e.g., Greenspan and Bona, 1993, *FASEB J.* 7:437-444; and Nissinoff, 1991, *J. Immunol.* 147:2429-2438). For example, antibodies  
15 that bind to a NHP domain and competitively inhibit the binding of a NHP to its cognate receptor can be used to generate anti-idiotypes that "mimic" the NHP and, therefore, bind and activate or neutralize a receptor. Such anti-idiotypic antibodies, or Fab fragments of such anti-idiotypes,  
20 can be used in therapeutic regimens involving a NHP signaling pathway.

Additionally given the high degree of relatedness of mammalian NHPs, NHP knock-out mice (having never seen the NHP, and thus never been tolerized to the NHP) have a unique  
25 utility, as they can be advantageously applied to the generation of antibodies against the disclosed mammalian NHPs (*i.e.*, a NHP will be immunogenic in NHP knock-out animals).

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended  
30 as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various

modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. All  
5 cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

## WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising at least 34 contiguous bases of nucleotide sequence from SEQ ID NO:1.  
5
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
  - (a) encodes the amino acid sequence shown in SEQ ID NO:2; and  
10
  - (b) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO:1 or the complement thereof.
- 15 3. An isolated nucleic acid molecule according to Claim 1 wherein said nucleotide sequence is a cDNA sequence.
4. An isolated nucleic acid molecule encoding a protein having the protease activity of the protease of SEQ ID NO:2, and that hybridizes to the nucleotide sequence of SEQ ID NO:1 under highly stringent conditions.  
20

## SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

<120> NOVEL HUMAN PROTEASE AND POLYNUCLEOTIDES  
ENCODING THE SAME

<130> LEX-0308-PCT

<150> US 60/270,320

<151> 2001-02-20

<160> 3

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1311

<212> DNA

<213> homo sapiens

<400> 1

```

atgtcctgct gtctggtctc accggtgggt gctccaggca tctgtgtatg cccctgtctg 60
tctggaccag gtgtgatcct aggagcgccc ctggcctcca gctgcgcagg agcctgtggt 120
accagcttcc cagatggcct caccctgag ggaaccagg cctccgggga caaggacatt 180
cctgcaatta accaagggt catcctggaa gaaacccag agagcagctt cctcatcgag 240
ggggacatca tccggccgag tcccttccga ctgctgtcag caaccagcaa caaatggccc 300
atgggtggta gtggtgtcgt ggaggtcccc ttcctgctct ccagcaagta cgatgagccc 360
agccgccagg tcatcctgga ggctcttgcg gagtgtgaac gttccacgtg catcagggtt 420
gtcacctatc aggaccagag agacttcatt tccatcatcc ccatgtatgg gtgcttctcg 480
agtgtggggc gcagtggagg gatgcagggt gtctccctgg cggccacgtg tctccagaag 540
ggccggggca ttgtccttca tgagctcatg catgtgctgg gcttctggca cgagcacacg 600
cgggcccagc gggaccgcta tatccgtgtc aactggaacg agatcctgcc aggcctttgaa 660
atcaacttca tcaagtctcg gaggcagcaac atgctgacgc cctatgacta ctctctgtgt 720
atgcactatg ggaggctcgc cttcagccgg cgtgggctgc ccaccatcac accactttgg 780
gccccagtg tccacatcgg ccagcgatgg aacctgagtg cctcggacat caccggggtc 840
ctcaaactct acggctgcag cccaagtggc cccaggcccc gtgggagagg gtcccatgcc 900
cacagcactg gtaggagccc cgctccggcc tccctatctc tgcagcggct tttggaggca 960
ctgtcggcgg aatccaggag cccgacccc agtgggttcca gtgcgggagg ccagcccgtt 1020
cctgcagggc ctggggagag cccacatggg tgggagtcct ctgccctgaa aaagctcagt 1080
gcagaggcct cggcaaggca gcctcagacc ctagcttcct cccaagatc aaggcctgga 1140
gcagggtgcc ccggtgttgc tcaggagcag tcctggctgg ccggagtgtc caccaagccc 1200
acagtcccat cttcagaagc aggaatccag ccagtccttg tccagggaag cccagctctg 1260
ccagggggct gtgtacctag aaatcatttc aaggggatgt ccgaagatta a 1311

```

<210> 2

<211> 436

<212> PRT

<213> homo sapiens

<220>

<221> VARIANT

<222> 227

<223> Xaa = Any Amino Acid

&lt;400&gt; 2

```

Met Ser Cys Cys Leu Val Ser Pro Val Gly Ala Pro Gly Ile Cys Val
 1           5           10           15
Cys Pro Cys Leu Ser Gly Pro Gly Val Ile Leu Gly Ala Pro Leu Ala
      20           25           30
Ser Ser Cys Ala Gly Ala Cys Gly Thr Ser Phe Pro Asp Gly Leu Thr
      35           40           45
Pro Glu Gly Thr Gln Ala Ser Gly Asp Lys Asp Ile Pro Ala Ile Asn
      50           55           60
Gln Gly Leu Ile Leu Glu Glu Thr Pro Glu Ser Ser Phe Leu Ile Glu
65           70           75           80
Gly Asp Ile Ile Arg Pro Ser Pro Phe Arg Leu Leu Ser Ala Thr Ser
      85           90           95
Asn Lys Trp Pro Met Gly Gly Ser Gly Val Val Glu Val Pro Phe Leu
      100          105          110
Leu Ser Ser Lys Tyr Asp Glu Pro Ser Arg Gln Val Ile Leu Glu Ala
      115          120          125
Leu Ala Glu Phe Glu Arg Ser Thr Cys Ile Arg Phe Val Thr Tyr Gln
      130          135          140
Asp Gln Arg Asp Phe Ile Ser Ile Ile Pro Met Tyr Gly Cys Phe Ser
145          150          155          160
Ser Val Gly Arg Ser Gly Gly Met Gln Val Val Ser Leu Ala Pro Thr
      165          170          175
Cys Leu Gln Lys Gly Arg Gly Ile Val Leu His Glu Leu Met His Val
      180          185          190
Leu Gly Phe Trp His Glu His Thr Arg Ala Asp Arg Asp Arg Tyr Ile
      195          200          205
Arg Val Asn Trp Asn Glu Ile Leu Pro Gly Phe Glu Ile Asn Phe Ile
      210          215          220
Lys Ser Xaa Ser Ser Asn Met Leu Thr Pro Tyr Asp Tyr Ser Ser Val
225          230          235          240
Met His Tyr Gly Arg Leu Ala Phe Ser Arg Arg Gly Leu Pro Thr Ile
      245          250          255
Thr Pro Leu Trp Ala Pro Ser Val His Ile Gly Gln Arg Trp Asn Leu
      260          265          270
Ser Ala Ser Asp Ile Thr Arg Val Leu Lys Leu Tyr Gly Cys Ser Pro
      275          280          285
Ser Gly Pro Arg Pro Arg Gly Arg Gly Ser His Ala His Ser Thr Gly
      290          295          300
Arg Ser Pro Ala Pro Ala Ser Leu Ser Leu Gln Arg Leu Leu Glu Ala
305          310          315          320
Leu Ser Ala Glu Ser Arg Ser Pro Asp Pro Ser Gly Ser Ser Ala Gly
      325          330          335
Gly Gln Pro Val Pro Ala Gly Pro Gly Glu Ser Pro His Gly Trp Glu
      340          345          350
Ser Pro Ala Leu Lys Lys Leu Ser Ala Glu Ala Ser Ala Arg Gln Pro
      355          360          365
Gln Thr Leu Ala Ser Ser Pro Arg Ser Arg Pro Gly Ala Gly Ala Pro
      370          375          380
Gly Val Ala Gln Glu Gln Ser Trp Leu Ala Gly Val Ser Thr Lys Pro
385          390          395          400
Thr Val Pro Ser Ser Glu Ala Gly Ile Gln Pro Val Pro Val Gln Gly
      405          410          415
Ser Pro Ala Leu Pro Gly Gly Cys Val Pro Arg Asn His Phe Lys Gly
      420          425          430
Met Ser Glu Asp

```

435

<210> 3  
 <211> 1586  
 <212> DNA  
 <213> homo sapiens

<400> 3  
 ccgaggctctg tccctgcctcc ttccttctctg cccctcctct acctcatagg tggggcacat 60  
 ggtccctttt ggtccccccta agggagctcc ttccttgagg tcatctagac cttggcacca 120  
 gttgggggttg agcagggagg ctgggaaggc tccttggtt tgtgctggag cctactcttc 180  
 ctagggactg agtcttaccg tctgatcccc cacaccacc ccatgtcctg ctgtctggtc 240  
 tcaccgggtg gtgctccagg catctgtgta tgccctgtc tgtctggacc aggtgtgatc 300  
 ctaggagcgc ccttggcctc cagctgcgca ggagcctgtg gtaccagctt cccagatggc 360  
 ctcacccttg agggaaacca ggcctccggg gacaaggaca ttcctgcaat taaccaaggg 420  
 ctcatccttg aagaaacccc agagagcagc ttcctcatcg agggggacat catccggccg 480  
 agtcccttcc gactgtgtg agcaaccagc aacaaatggc ccatgggtgg tagtggtgtc 540  
 gtggaggtcc ccttctgtg ctccagcaag tacgatgagc ccagccgcca ggtcatcctg 600  
 gaggctcttg cggagtttga acgttccacg tgcacaggt ttgtcaccta tcaggaccag 660  
 agagacttca tttccatcat ccccatgtat ggggtgttct cgagtgtggg gcgcagtgga 720  
 gggatgcagg tggctctcct ggcgcccacg tgtctccaga agggccgggg cattgtcctt 780  
 catgagctca tgcattgtgct gggcttcttg cacgagcaca cgcgggcca cggggaccgc 840  
 tatatccgtg tcaactggaa cgagatcctg ccaggctttg aaatcaactt catcaagtct 900  
 cggagcagca acatgctgac gccctatgac tactcctctg tgatgcacta tgggaggctc 960  
 gccttcagcc ggcgtgggct gccaccatc acaccacttt gggcccccag tgtccacatc 1020  
 ggccagcgat ggaacctgag tgccctggac atcaccggg tcctcaaact ctacggctgc 1080  
 agcccaagtg gccccaggcc ccgtgggaga ggggtccatg cccacagcac tggtaggagc 1140  
 cccgctccgg cctccctatc tctgcagcgg cttttggagg cactgtcggc ggaatccagg 1200  
 agccccgacc ccagtgggtc cagtgcggga ggcagcccg ttcctgcagg gcctggggag 1260  
 agcccacatg ggtgggagtc ccctgccctg aaaaagctca gtgcagaggg ctcggaagg 1320  
 cagcctcaga ccctagcttc ctccccaaga tcaaggcctg gagcaggtgc ccccggtgtt 1380  
 gctcaggagc agtcctggct ggccggagtg tccaccaagc ccacagtccc atcttcagaa 1440  
 gcaggaatcc agccagtccc tgtccaggga agcccagctc tgccaggggg ctgtgtacct 1500  
 agaaatcatt tcaaggggat gtccgaagat taagcctgtg gcttctgtcc ccaagtaggg 1560  
 agggcatcct ctgcccagtg gagctg 1586